9-2013

Web-scale near-duplicate search: Techniques and applications

Chong-wah NGO
Singapore Management University, cwngo@smu.edu.sg

Changsheng XU

Wessel KRAAIJ

Abdulmotaleb EL SADDIK

Web-Scale Near-Duplicate Search: Techniques and Applications

Chong-Wah Ngo
City University of Hong Kong

Changsheng Xu
Chinese Academy of Sciences

Wessel Kraaij
TNO and Radboud University Nijmegen

Abdulmotaleb El Saddik
University of Ottawa

As the bandwidth accessible to average users has increased, audiovisual material has become the fastest growing datatype on the Internet. The impressive growth of the social Web, where users can exchange user-generated content, contributes to the overwhelming number of multimedia files available. Among these huge volumes of data, a large numbers of near duplicates and copies exist. File copies are easy to detect using hashes. However, near duplicates are based on the same original content but have been edited and postprocessed, resulting in different files. Another type of near duplicate includes footage of the same event or scene.

Detecting near duplicates poses a challenge for multimedia content analysis, especially when speed, scale, and copied fragment length are pushed to operational levels. Near duplicates carry both informative and redundant signals, for example, providing rich visual clues for indexing and summarizing videos from different sources. Still, the excessive amount of near duplicates streamed over Internet demands scalable techniques for copyright infringement detection, advertisement tracking, and content monitoring for forensic applications. As a result, there is strong interest from industry, academia, and governmental agencies in Web-scale search, elimination, detection, and use of near duplicates for various multimedia applications.

This special issue presents some of the most recent advances in the research on Web-scale near-duplicate search and explores the potential for bringing this research a substantial step further. It contains high-quality contributions addressing various aspects of the Web-scale near-duplicate search problem in a number of relevant domains.

In This Issue

This special issue features five articles addressing the fundamental issues of near-duplicate retrieval as well as their limitations and applicability for emerging multimedia applications. The topics range from feature representation, matching, and indexing from different novel aspects to the adaptation of current technologies for mobile media search and photo archaeology mining.

Partial duplicate retrieval is one of the topics receiving intensive research attention because of its potential applications for product search. The difficulties comes from the fact that users often have a search target or object in mind, and the similarity comparison often needs to be conducted on the region rather than the full-image level. ''Partial-Duplicate Image Retrieval via Saliency-Guided Visual Matching'' by Liang Li, Shuqiang Jiang, Zheng-Jun Zha, Zhipeng Wu, and Qingming Huang addresses this challenging issue by proposing the removal of image background noise from comparison by visual saliency modeling. In this way, feature indexing and matching only concentrates on visually salient regions, which more likely correspond to the user search targets. A beauty of this method is that the saliency values can be elegantly leveraged as spatial constraints for similarity measure and efficiently indexed into a two-level inverted file structure for scalable search of partial duplicate images.

1070-986X/13/$31.00 © 2013 IEEE

Published by the IEEE Computer Society

Although there is rich literature on indexing and searching similar images, the overhead imposed on memory for trading higher search efficiency is often overlooked. ''Web-Scale Image Retrieval Using Compact Tensor Aggregation of Visual Descriptors'' by Romain Negrel, David Picard, and Philippe-Henri Gosselin presents an overview of the existing visual descriptors and their associated indexing techniques using hashing, bag-of-words, and tree-based representation. This article addresses the problem of memory consumption by proposing a practical way of generating highly compact signatures that involves techniques in tensor aggregation, principal component analysis (PCA), and kernel PCA at different stages. The method contributes to the state of the art by improving the family of Fisher vector descriptors in terms of the feature discriminative power and the size of feature signature. On a million-scale dataset, the method only needs 61 Mbytes of memory, yet still maintains a comparable and sometimes even higher performance than the similar types of techniques.

Similar in spirit, ''Nested-SIFT for Efficient Image Matching and Retrieval'' by Pengfei Xu, Lei Zhang, Kuiyuan Yang, and Hongxun Yao takes a deeper look at scale-invariant feature transform (SIFT) descriptors, the state-of-the-art visual descriptors. The authors propose an elegant way of constructing SIFT groups, embedding geometric information, and eventually compacting a group as a 64-bit binary signature named nested-SIFT for near-duplicate detection. Nested-SIFT utilizes the nested relationship of SIFT descriptors and naturally groups the local interest points of different scales to generate feature signature. Nested-SIFT is more discriminative for embedding spatial information, and its compact version, generated using SimHash, is more efficient in visual matching. The experimental results show that Nested-SIFT can nicely optimize different performance parameters (such as retrieval accuracy, memory cost, and search speed) of a million-scale image retrieval system.

Despite various efforts dedicated to packing visual descriptors into tiny signatures, not all techniques can practically run on resource-constraint platforms such as mobile devices, which have limited memory space and computing power. For example, the projection matrix used by some techniques to map high-dimension feature vectors to a low-dimensional space simply cannot fit into the memory of mobile devices. Using machine-learning techniques, ''Scalable Mobile Video Retrieval with Sparse Projection Learning and Pseudo Label Mining'' by Guan-Long Wu, Yin-Hsi Kuo, Tzu-Hsuan Chiu, Winston H. Hsu, and Lexing Xie proposes an approach to learning sparse projection matrices that is feasible to load on mobile platforms. The learning can optionally take external information such as knowledge from Wikipedia and text snippets from Google search results as input, creating a semantically aware projection and producing compact signatures meeting the constraints of mobile media retrieval.

Lastly, ''Large-Scale Image Phylogeny: Tracing Image Ancestral Relationships'' by Zanoni Dias, Siome Goldenstein, and Anderson Rocha demonstrates the use of near-duplicate detection techniques for an interesting problem called *multimedia phylogeny*, which aims to explore the history and evolutionary path of media objects. The article shows how large-scale archaeology relationship mining can be realized by the partial construction of dissimilarity matrices for near-duplicate measurement. Based on this technique, different ways of building a phylogeny tree for visualizing the casual and ancestral relationships of images are presented. This article uses a real example on a famous photo, ''The Situation Room'' released by the White House, to illustrate the ''evolutionary process'' of the photo on different social media websites.

Future Directions

Data-driven technology has been a hot term for some time. The philosophy is that traditionally hard problems in recognition can become easy if there is a huge data pool behind them that can turn the recognition into a search problem. Techniques in Web-scale near duplicate search play an important part in making this philosophy a practical engineering task for multimedia applications. Today, we have seen examples, such as leveraging near-duplicate images for location estimation, 3D reconstruction, product search, and media annotation. We envision that more examples will emerge, and near-duplicate search is becoming a ''standard step'' in processing large multimedia archives, such as shot boundary detection in structuring video content and as a ''preprocessing step''

for Web-scale Big Data analysis. In addition, we also envision that more tailor-made technologies will come to address the needs of various emerging multimedia applications such as forensics, memory- and power-efficient algorithms for meeting computing constraints on mobile devices, and the wise use of partial-duplicate information in different search scenarios for online commercial product search, monitoring, and advertisement. **MM**

Chong-Wah Ngo is an associate professor at the City University of Hong Kong and the founding leader of the VIREO Research Group. His research interests include large-scale multimedia information retrieval, data mining, indexing, and visualization. Ngo has a PhD in computer science from the Hong Kong University of Science and Technology. Contact him at cscwngo@cityu.edu.hk.

Chengsheng Xu is a professor in the National Lab of Pattern Recognition, Institute of Automation, at the Chinese Academy of Sciences and executive director of the China-Singapore Institute of Digital Media. His research interests include multimedia content analysis, indexing, and retrieval; pattern recognition; and computer vision. Xu has a PhD from Tsinghua University, China. Contact him at csxu@nlpr.ia.ac.cn.

Wessel Kraaij is a senior scientist and leader of the Media Mining Group at TNO (Netherlands Organization for Applied Scientific Research) and a part-time full professor at the Radboud University Nijmegen. His research interests include multimedia information retrieval, behavioral analytics, and context modeling. He is one of the coordinators of the TRECVID Benchmark Conference. Kraaij has a PhD in computer science from the University of Twente. Contact him at kraaijw@acm.org.

Abdulmotaleb El Saddik is a university research chair and professor in the School of Electrical Engineering and Computer Science and the director of the Multimedia Communications Research Laboratory (MCRLab) at the University of Ottawa. His research interests include the knowledge and understanding of multimedia computing, communications, and applications, particularly in the digitization, communication, and security of the sense of touch, or haptics. El Saddik has a PhD from Darmstadt University of Technology, Germany. He is a fellow of IEEE. Contact him at elsaddik@uottawa.ca.